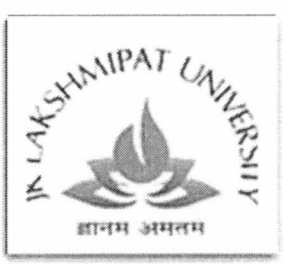
Data Mining- an Approach to Identify Electricity Waste patterns in Engineering Colleges

Research Proposal submitted in partial fulfillment of the requirements for the Degree of

DOCTOR OF PHILOSOPHY
in
ENGINEERING

By
Ram Lal Yadav

Under the Guidance of
Dr. Sonal Jain
Associate Professor



JK Lakshmipat University
(Recognized by Govt. of Rajasthan and covered u/s 2(f) of the UGC Act 1956)
Laliya Ka Vas, P.O. Mahapura, Ajmer Road, Jaipur - 302 026 (Rajasthan) India

OUTLINE OF THE PROPOSED TOPIC OF RESEARCH

Name of the Candidate	:	Ram Lal Yadav
ID No.	:	2012PhDEngg007
Place of Research Work and Organization	:	JKLU, Jaipur
Proposed Supervisor Name	:	Dr. Sonal Jain
Qualification	:	Ph. D.
Designation	:	Associate Professor
Organization	:	JKLU, Jaipur

Proposed Topic of Research

Data Mining- an Approach to Identify Electricity Waste patterns in Engineering Colleges

Objectives of the Proposed Research

A significant portion of the energy consumption in engineering institutes is consumed by classrooms lighting and fans. The occupancy patterns in engineering institutes are not stable and predictable, and thus, alternative solutions may be required to match energy consumption and occupancy in order to increase energy efficiency. The proposed methodology is explained briefly, first the data will be collected and pre-processed in order to reduce numerosity, noise and create higher-level concepts. This will results in better and more meaningful rules extracted from the dataset. Second, an energy assessment is conducted in order to evaluate the current lighting-energy performance of the space under study. Third, data mining algorithm will be applied to the resulting database in order to discover knowledge and relationships between variables of the dataset with a focus on discovering the pattern of energy waste, and simulation of the implementation of some simple remedies based on the discovered rules and quantify the potential energy saving.

A data mining tool, will be applied to the data in order to extract association rules and explore lighting waste patterns.

Background of the Proposed Research

Introduction:

Over the past years, there have been significant efforts to reduce lighting energy waste in the residential and commercial / institutional sectors. One of the effective approaches, particularly for commercial/institutional sectors, has been occupancy based lighting control systems [1,3,5,6]. The basic idea behind these systems is turning the lights off when a state of no occupancy is detected in the monitored space.

With a projected 70% of the global population living in urban areas by 2050, the United Nations Environment Programme has underlined the need to promote a low carbon, resource efficient and socially inclusive 'green' economy [1,2]. In particular, energy efficient buildings that reduce both emissions and waste are of significant interests. One target area to increase the energy efficiency of buildings and reduce waste is through investments in more efficient lighting fixtures and the use of time-scheduled switching or occupancy sensors. Lighting alone accounted for 19% of total electricity consumption in the world in the year 2005, generating 1900 million tonnes of CO2 emissions [3].

Association Rule Mining (ARM) a data mining approach is a popular and powerful tool for discovering interesting relationships between the attributes of a dataset. ARM has been applied to many areas, including marketing promotion, Web usage mining, intrusion detection and bioinformatics. ARM has become one of the most important research areas in data mining [7]. Current studies on association rule mining focus on finding Boolean/ quantitative association rules from certain databases or Boolean association rules from probabilistic databases. So according to the study of various research papers, ARM tool can be used in our problem

However, a substantial number of datasets in real-world applications are actually uncertain in nature because imprecision usually exists in the process of collecting, transmitting, storing and preprocessing data. For example, because of the dynamicity of our environment and their own limited ability, sensors measure and collect data only in an imprecise way. In many situations, to preserve private information, we might intentionally change certain data into uncertain data. In such databases, the existence of quantitative attribute values in a transaction is best captured by a likelihood measure or probability because of the observation

that the generated data are inherently uncertain. In this light, ARM in an uncertain database has recently attracted much attention in the field of data mining [7].

Literature Review of Research Topic

Data association mining for identifying lighting energy waste patterns in educational institutes

David F. Motta Cabrera et. al., 2013 exploits the efficiency of Association Rule Mining (ARM) an experimental research on quantifying and understanding lighting energy waste patterns in a post-secondary educational institute. Data has been collected over a full academic year in three typical classrooms. Data association mining, a powerful data mining tool, is applied to the data in order to extract association rules and explore lighting waste patterns. The simulations results show that if the waste patterns are avoided, significant savings, as high as 70% of the current energy use, are achievable.

Exploring Sequential and Association Rule Mining for Pattern-based Energy Demand Characterization

Leneve Ong et. al., 2013 presented the relationship between occupant activity and electricity consumption is inextricably linked. It has been difficult to both gather detailed energy data and information about occupants' daily lives as well as understand their relationship quantitatively. There is significant past work on activity recognition in homes and load prediction, but there is limited understanding of how activities can inform consumption or vice versa. Our work begins by characterizing power data as provided by plug-level meters from one household. Association and sequential rule mining techniques are applied to extract explicit rules that may be useful for forming the basis of demand patterns. Initial findings include the identification of device groups but highlight the challenges of modeling complex patterns and event rarity.Authors suggest that further this algorithm is used in similar classification problems with large data sets of nasopharyngeal carcinoma and lung cancer.

Analysis of energy use at university Campus

Aleksandra Sretenovic, 2013 presented in his thesis that the study of the building energy demand has become a topic of great importance, because of the significant increase of interest in energy sustainability. University campuses represent specific groups of diverse

buildings, with significant energy consumption. They consist of many different buildings, representing small-scale town for itself. Therefore, they provide an excellent testbed to characterize and understand energy consumption of group of "mixed use" buildings. Suitable building database for University campus NTNU Gloshaugen is created, and available data of heating and electricity energy use are collected and organized. Having correct and reliable data is essential, so data error analysis using statistical methods is performed. Heating energy use was modeled using Matlab statistical toolbox functions. Creating a model of energy use helps in future building planning; it can provide useful information about most probable energy consumption for similar buildings, or predict energy use in different conditions.

Using Rule Mining to Understand Appliance Energy Consumption Patterns

Sami Rollins et. al., Abstract—Managing energy in the home is key to creating a sustainable future for our society. More tools are increasingly available to measure home energy usage, however these tools provide little insight into questions such as why an appliance consumes more energy than normal or what kinds of behavioural changes might be most likely to reduce energy usage in the home. To answer these questions, a deeper understanding of the causal factors that influence energy usage is necessary. In this work, we conduct a broad study of factors that influence energy consumption of individual devices in the home. Our first contribution is collection of a context-rich data set from six homes across the United States. The second contribution of this work is a set of insights into key factors influencing energy usage derived by the novel application of a rule mining algorithm to identify significant associations between energy usage and four key features: hour of the day, day of the week, use of other appliances in the home, and user-supplied annotations of activities such as working or cooking. Our analysis confirms our hypothesis that, though most devices show a regular pattern of daily or weekly use, this is not true for all devices. Associations that relate use of two different devices in the same home are often stronger, and are observed for nearly 25% of device uses. Overall, we observe that the associations derived from the first five weeks of data in our data set are sufficient to explain nearly 70% of the device uses in the subsequent five weeks of data, and over 90% of the associations identified during the first five weeks recur in the latter portion of the data set. The associations identified by our approach may be used to aid in end-user applications that heighten awareness and encourage energy savings, improve energy disaggregation algorithms, or even detect anomalous uses that may signal problems in aging-in-place homes.

In this work, they undertake a study of the causal factors that impact energy usage at the granularity of individual appliances in the home.

Fault Detection Analysis of Building Energy Consumption Using Data Mining Techniques

Imran Khan et. al, 2013 This study describes three different data mining techniques for detecting abnormal lighting energy consumption using hourly recorded energy consumption and peak demand (maximum power) data. Two outliers' detection methods are applied to each class and cluster for detecting abnormal consumption in the same data set. In each class and cluster with anomalous consumption the amount of variation from normal is determined using modified standard scores. The study will be helpful for building energy management systems to reduce operating cost and time by not having to detect faults manually or diagnose false warnings. In addition, it will be useful for developing fault detection and diagnosis model for the whole building energy consumption.

Energy consumption of both residential and commercial buildings has steadily increased, reaching figures up to 40% in developed countries. Increasing demand for building services and high thermal comfort levels, together with the amount of time spent indoors, will increase the energy demand in future. There is an increasing realization that many buildings do not perform as intended by their designers. Typical buildings consume 20% more energy than necessary due to faulty construction, malfunctioning equipment, incorrectly configured control systems and inappropriate operating procedures.

CAR-Miner: An efficient algorithm for mining class-association rules

Loan T.T. Nguyen et. al., 2013 Building a high accuracy classifier for classification is a problem in real applications. One high accuracy classifier used for this purpose is based on association rules. In the past, some researchers showed that classification based on association rules (or class-association rules – CARs) has higher accuracy than that of other rule-based methods such as ILA and C4.5. However, mining CARs consumes more time because it mines a complete rule set. Therefore, improving the execution time for mining CARs is one of the main problems with this method that needs to be solved. In this paper, we propose a new method for mining class-association rule. Firstly, we design a tree structure for the storage frequent itemsets of datasets. Some theorems for pruning nodes and computing information in the tree are developed after that, and then, based on the theorems, we propose an efficient algorithm for mining CARs. Experimental results show that our approach is more

efficient than those used previously. For association rule mining, the target attribute (or class attribute) is not pre-determined. However, the target attribute must be pre-determined in classification problems. Thus, some algorithms for mining classification rules based on association rule mining have been proposed.

FARP: Mining fuzzy association rules from a probabilistic quantitative database

Bin Pei et. al., 2013 says that the current studies on association rule mining focus on finding Boolean/quantitative association rules from certain databases or Boolean association rules from probabilistic databases. However, little work on mining association rules from probabilistic quantitative databases has been mentioned because the simultaneous measurement of quantitative information and probability is difficult. By introducing anovel Shannon-like Entropy, we aggregate and measure the information contained in an item with the coexistence of fuzzy uncertainty hidden in quantitative values and random uncertainty. We then propose Support and Confidence metrics for a fuzzy–probabilistic database to quantify association rules. Finally, we design an algorithm, called FARP (mining Fuzzy Association Rules from a Probabilistic quantitative data),to discover frequent fuzzy–probabilistic itemsets and fuzzy association rules using the proposed interest measures. The experimental results show the effectiveness of our method and its practicality in real-world applications.

According to Bin Pei et. al., 2013 Association rule mining (ARM)is a popular and well-stud ied tool for discovering interesting relationship s between the attributes of a dataset. ARM has been applied to many areas, including marketing promotion, Web usage mining, intrusion detection and bioinformatics. ARM has become one of the most important research areas in data mining. In the Apriori algorithm for determining association relationships hidden in Boolean valued transaction databases. A Boolean association rule can be expressed as X□Y, where X and Y are sets of Boolean attributes, indicating that an object that satisfies attributes X might also satisfy attributes Y. This specific problem is termed as Boolean ARM from a certain database because the items that occur in a transaction are known for certain. However, a substantial number of datasets in real-world applications are actually uncertain in nature because imprecision usually exists in the process of collecting, transmitting, storing and preprocessing data. For example, because of the dynamicity of our environment and their own limited ability, sensors measure and collect data only in an imprecise way.

Association Rules Mining: A Recent Overview

Sotiris Kotsiantis et. al., 2006 In this paper, they provide the preliminaries of basic concepts about association rule mining and survey the list of existing association rule mining techniques. Of course, a single article cannot be a complete review of all the algorithms, yet we hope that the references cited will cover the major theoretical issues, guiding the researcher in interesting research directions that have yet to be explored. Association rule mining, one of the most important and well researched techniques of data mining, was first introduced in [1]. It aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories. Association rules are widely used in various areas such as telecommunication networks, market and risk management, inventory control etc.

Analysis of Sampling Techniques for Association Rule Mining

Venkatesan T. Chakaravarthy et. al., 2009 they present a comprehensive theoretical analysis of the sampling technique for the association rule mining problem. Most of the previous works have concentrated only on the empirical evaluation of the effectiveness of sampling for the step of finding frequent itemsets. To the best of our knowledge, a theoretical framework to analyze the quality of the solutions obtained by sampling has not been studied. Our contributions are two-fold. First, we present the notions of ^{2}-close frequent itemset mining and ^{2}-close association rule mining that help assess the quality of the solutions obtained by sampling. Secondly, we show that both the frequent items mining and association rule mining problems can be solved satisfactorily with a sample size that is independent of both the number of transactions size and the number of items.

Deriving Private Information from Association Rule Mining Results

Zutao Zhu et. Al., Data publishing can provide enormous benefits to the society. However, due to privacy concerns, data cannot be published in their original forms. Two types of data publishing can address the privacy issue: one is to publish the sanitized version of the original data, and the other is to publish the aggregate information from the original data, such as data mining results. There have been extensive studies to understand the privacy consequence in the first approach, but there is not much investigation on the privacy consequence of publishing data mining results, although, it is well believed that publishing data mining results can lead to the disclosure of private information. We propose a systematic method to study the privacy consequence of data mining results. Based on a well-established theory, the

principle of maximum entropy, we have developed a method to precisely quantify the privacy risk when data mining results are published.

Methodology

The Study

The main purpose of the research study is to explore the previous methodology in the direction of energy waste pattern detection and find the advantages and disadvantages so that we can make a better framework where efficient energy waste pattern detection technique can be developed, so that electric power waste can be detected, and it could be saved. After studying several research papers we observe that there is lot of work in this area. But there is still the need of betterment. So our direction of research is to the improvement in terms of electric power waste detection using Data Association Mining Technique.

The Design

Our Working procedure can be better explained by the following flowchart shown in figure 1.

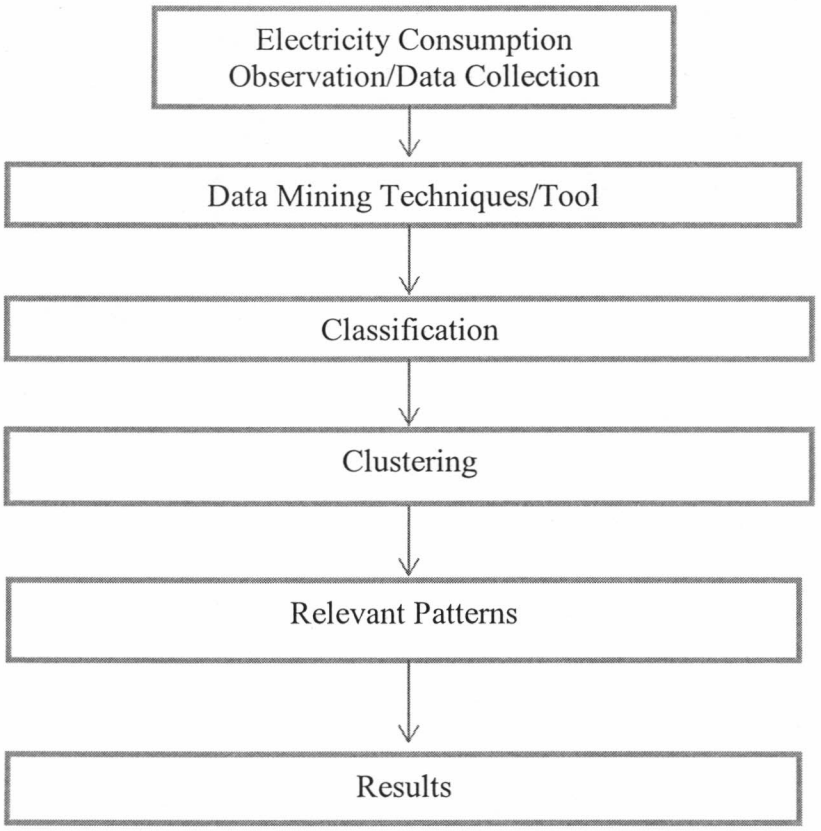


Figure 1: Process Flowchart

In this process we use the previous research dataset and will observe the electricity consumption to find better identification of energy waste patterns. First find support, confidence and lift values then we will apply the association classification algorithm. After this apply clustering algorithm for creating clusters. Then relevant patterns would be observed. Based on the relevant patterns we will get the better results to identify energy waste patterns.

So our procedure is further divided into six parts:

1) Electricity Consumption Observation/Data Collection
2) Data Mining Techniques/Tools
3) Classification
4) Clustering
5) Relevant Patterns
6) Results

Selection of Data Set (Sample)

Sub-meters/sensors would be installed in the institution rooms, then units/consumption will be observed throughout the ear (in all rain, winter and summer season).

Association Rule Classification

Although association rules have been predominantly used for data exploration and description, the interest in using them for prediction has rapidly increased in the data mining community. In order to mine only rules that can be used for classification, we can use association rule mining algorithm like, Apriori to handle user-defined input set constraints. It may classify the rules based on support, confidence and lift.

Frequent Itemset Mining (FIM) is considered to be one of the elemental data mining problems that intends to discover groups of items or values or patterns that co-occur frequently in a dataset. It is of vital significance in a variety of Data Mining tasks that aim to mine interesting patterns from databases, like association rules, correlations, sequences, episodes, classifiers, clusters and the like. The proposed approach utilizes an efficient algorithm called Association Rule Mining Algorithm. The proposed algorithm is employed

for the extraction of association rules from the clustered dataset besides performing efficiently when the database consists of very long item sets specifically.

Support count: The support count of an itemset X, denoted by X.count, in a data set T is the number of transactions in T that contain X. Assume T has n transactions (Agrawal, R. et al. , 1993)

$$support = \frac{(X \cup Y).count}{n}$$

Confidence: The rule holds in T with confidence conf if conf% of transactions that contain X also contain Y (Agrawal, R. et al. , 1993)

$$confidence = \frac{(X \cup Y).count}{X.count}$$

Clustering Algorithm

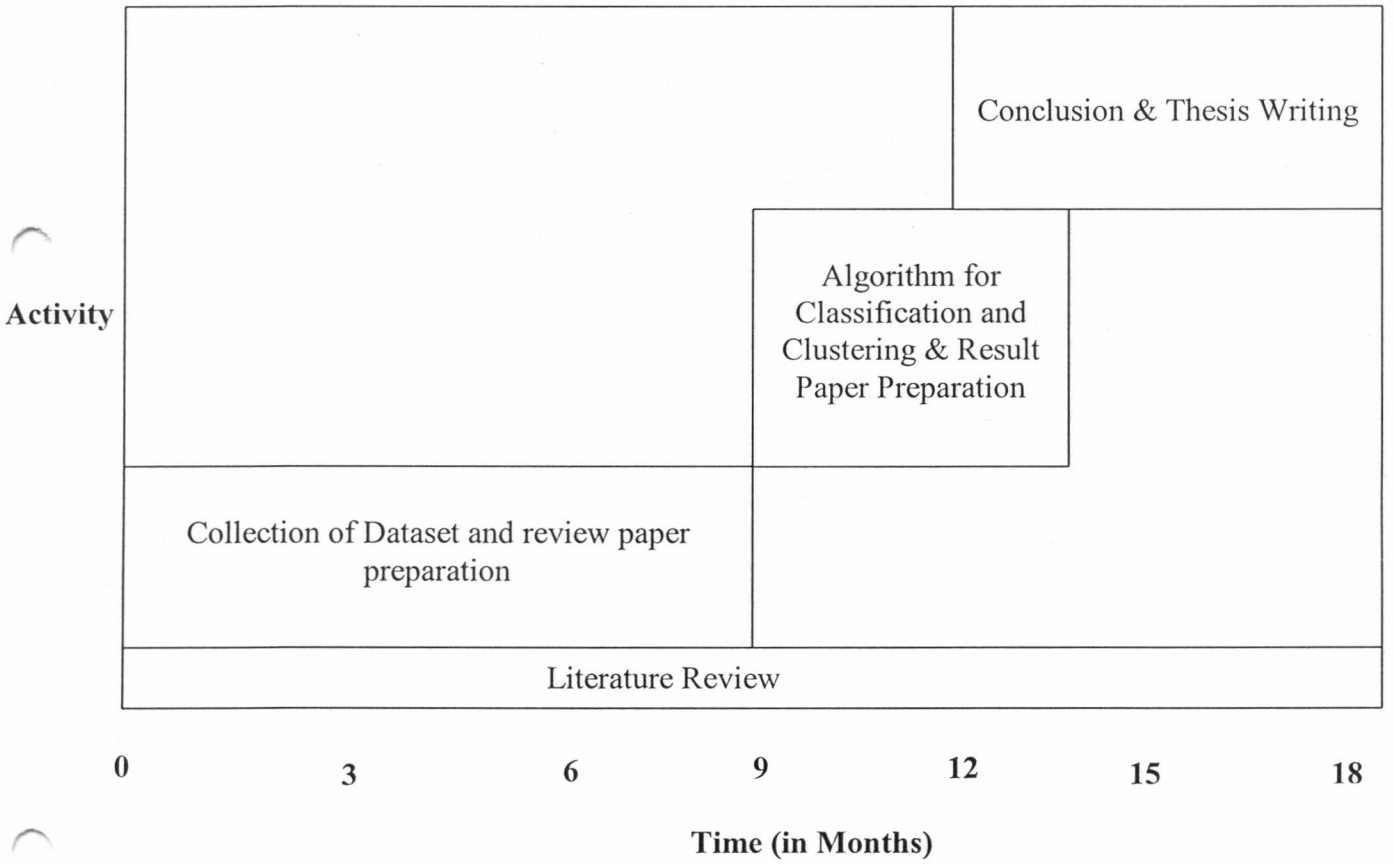
The categorization of objects into various groups or the partitioning of data set into subsets so that the data in each of the subset share a general feature, frequently the proximity with regard to some defined distance measure, is known as Clustering. The clustering problem has been addressed in numerous contexts besides being proven beneficial in many applications. Clustering medical data into small yet meaningful clusters can aid in the discovery of patterns by supporting the extraction of numerous appropriate features from each of the clusters thereby introducing structure into the data and aiding the application of conventional data mining techniques. Numerous methods are available in the literature for clustering. I may employ the renowned K-Means clustering algorithm in our approach.

The k-means algorithm is one of the widely recognized clustering tools that are applied in a variety of scientific and industrial applications. K-means groups the data in accordance with their characteristic values into K distinct clusters. Data categorized into the same cluster have identical feature values. K, the positive integer denoting the number of clusters, needs to be provided in advance.

Work Plan

The work-plan for the above-mentioned six activities is shown in figure 2.

Figure 2: WORK PLAN



REFERENCES

[1] David F. Motta Cabrera, Hamidreza Zareipour (2013), *"Data association mining for identifying lighting energy waste patterns in educational institutes"*, Energy and Buildings 62 (2013) 210–216.

[2] UNEP (2011), *"Towards a Green Economy: Pathways to Sustainable Development and Poverty Eradication – A Synthesis for Policy Makers"*.

[3] International Energy Agency (2006), *"Light's Labour Lost: Policies for Energy-Efficient Lighting"*.

[4] America's Energy Future Panel on Energy Efficiency Technologies (2010), *"Real Prospects for Energy Efficiency in the United States"*, Tech. Rep. USA.

[5] V. Garg, N. Bansal (2000), *"Smart occupancy sensors to reduce energy consumption"*, Energy and Buildings 32 (1) (2000) 81–87.

[6] A.D. Galasiu, G.R. Newsham, C. Suvagau, D.M. Sander (2007), *"Energy saving lighting control systems for open-plan offices: a field study"*, Tech. Rep., Institute for Research in Construction, Canada, URL: http://www.nrc-cnrc.gc.ca/obj/irc/doc/pubs/nrcc49498/nrcc49498.pdf.

[7] Bin Pei, Suyun Zhao, Hong Chen, Xuan Zhou, Dingjie Chen (2013), *"FARP: Mining fuzzy association rules from a probabilistic quantitative database"*, Information Sciences 237 (2013) 242–260.

[8] T. Moore, D. Carter, A. Slater (2003), *"Long-term patterns of use of occupant controlled office lighting"*, Lighting Research and Technology 35 (1) 43–57.

[9] Imran Khan, Alfonso Capozzoli, Stefano Paolo Corgnati, Tania Cerquitelli (2013), *"Fault Detection Analysis of Building Energy Consumption Using Data Mining Techniques"*, Energy Procedia 42 (2013) 557 – 566.

[10] Loan T.T. Nguyen, Bay Vo, Tzung-Pei Hong, Hoang Chi Thanh, *"CAR-Miner: An efficient algorithm for mining class-association rules"*, Expert Systems with Applications 40 (2013) 2305–2311.

[11] Venkatesan T.Chakaravarthy, Vinayaka Pandit, Yogish Sabharwal (2009), *"Analysis of Sampling Techniques for Association Rule Mining"*, ACM 978-1-60558-423-2/09/0003.

[12] Sotiris Kotsiantis, Dimitris Kanellopoulos (2006), *"Association Rules Mining: A Recent Overview"*, GESTS, Vol.32 (1), 2006, pp. 71-82.

[13] Zutao Zhu, Guan Wang, Wenliang Du, *"Deriving Private Information from Association Rule Mining Results"*, Syracuse University, Syracuse, NY, USA 13244.

[14] Yo-Ping Huang, Li-Jen Kao, Frode-Eika Sandnes, *"A Prefix Tree-based Model for Mining Association Rules from Quantitative Temporal Data"*.

[15] Azzedine Boukerche and Samer Samarah (2007), *"An Efficient Data Extraction Mechanism for Mining Association Rules from Wireless Sensor Networks"*, 1-4244-0353-7/07 ©2007 IEEE.

[16] Duc H. Tran, Wee Keong Ng, Wei Zha (2009), *"CRYPPAR: An Efficient Framework for Privacy Preserving Association Rule Mining over Vertically Partitioned Data"*, 978–1–4244–4547–9/09 ©2009 IEEE.

[17] XIN-YE LI, JIN-SHA YUAN, YING-HUI KONG (2007), *" MINING ASSOCIATION RULES FROM XML DATA WITH INDEX TABLE"*, 1-4244-0973-X/07 ©2007 IEEE.

[18] Sussan Einakian, M. Ghanbari (2006), *"Parallel Implementation of Association Rule in Data Mining"*, 0-7803-9457-7/06 ©2006 IEEE.

[19] Qin Ding, Qiang Ding, and William Perrizo (2008), *"PARM-An Efficient Algorithm to Mine Association Rules From Spatial Data"*, 10.1109/TSMCB.2008.927730.

[20] W. Sammouri, E. Come, L. Oukhellou, P. Aknin, Ch-E. Fonlladosa, and K. Prendergast (2012), *"Temporal association rule mining for the preventive diagnosis of onboard subsystems within floating train data framework"*, 978-1-4673-3063-3/12 ©2012 IEEE.

[21] Rakesh Agrawal, Ramakrishnan Srikant (1994), *"Fast Algorithms for Mining Association Rules"*, Proceedings of the 20th VLDB Conference Santiago, Chile, 1994.